## The Discovery of Handwashing – A Python Recreation

**Description**

The age-old saying 'waste not, want not' resonates profoundly in the current environment of heightened global awareness and concern towards environmental issues. His discoveries resulted in a sharp declination in patients’ mortality rate and led to the development of sanitary measures still in use in healthcare settings to this day. In this project, we are on a mission to replicate Dr. Semmelweis's discovery where Python will be the tool of choice, a dominating language in the fields of data analytics and visualization.

The project is based on historical data sets representing various groups of data which have mortality rates in a hospital that occurred during different periods. We are going to employ Python to diligently examine the datasets in detail. We will be exploring the connection on the relationship between the variables and reveal the impact of hand washing on mortality rates. Through Python, we will boost data manipulation and statistical analysis to perform calculations such as mortality rates prior and after the adoption of handwashing protocols.

Data Loading and Exploration: Our data analysis will commence with loading the historical dataset into the Python environment and conducting a preceding exploratory analysis to get the idea of the dataset structure and content.

Relationship Analysis: By using Python's data analysis tools, we will be able to get a deeper understanding and relations between variables like the time period, deaths rates, and handwashing methods.

Calculations and Comparisons: Working out the Python numerical computing element, we are going to calculate mortality rates before and after employing handwashing protocols. The purposes of this step are to quantify how much handwashing influence on decreasing mortality rates.

Visualization: A crucial tool for the delivery of insights through visual representation of data is thus to be highlighted. There will be Python visualization libraries to be employed in order to generate graphs and charts that will be easy to interpret and will effectively show the effect of hand washing on the mortality rates over time.

**Analysis of the relationship between variables over different periods**

This project intends to analyze the connection between several variables that are time-dependent so that we can grasp the role and the effect of the hand washing on the mortality rates. On this based on the works of Dr. Ignaz Semmelweis. To do this, we shall start by getting a historical data that indicates the mortality rates in a hospital setting separate by different periods.

Our analysis will include look at shared parameters, like if was a specific period or just the understanding of what will reduce death rate, handwashing or simply we need a well-trained nurse. Python is data analysis tool to analyze patterns and we'll explore it into the data to do so. We will research if a significant change in mortality incidences would occur not long after putting the handwashing protocol into enforcement.

To numerically evaluate the performance of the hand washing, we will do arithmetic calculations according to the change in the before-and-after mortality rates after the introduction of hand hygiene practices. This stage will involve the use of statistical analysis to identify any apparent differences in mortality rate between the periods of hand washing being catered for and the period where hand washing was not being catered for.

Apart from just the visualization of our results, the visual form of representation will have the appropriate role in the presentation of our discoveries. We will visualize heat maps, whereas there are relationships between variables over different period which are graphical outputs created through Python's visualization libraries. These visual representations including among others, charts will be presented to give a clear picture of what people can do to help reduce the global mortality rate by hand washing and the message will be served in the person’s knowledge.

The purpose of this research is to reveal the far-reaching effect of personal hygiene on health. Additionally, it draws a parallel between the modern world and the past when data monitoring, analysis, and statistics played a vital role in discovering medical breakthroughs.

**Now Let’s Look at the Code Snippets**

**First We will import all the necessary module libraries :**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

Now to read the yearly dataset that is provided to us

yearly\_df = pd.read\_csv("yearly\_deaths\_by\_clinic.csv")

yearly\_df

yearly\_df.shape

This code provides us with the details of the dataset for example number of columns and number of rows present :

yearly\_df.info()

This gives all the information of the dataset such as datatype of rows and , NULL or NON NULL

yearly\_df.groupby("clinic") ["deaths"].sum()

Grouping the data by the "clinic" column, and then calculates the sum of the "deaths" column for each group.

yearly\_df["Proportion of Deaths"] = yearly\_df["deaths"] / yearly\_df["births"]

yearly\_df

calculating the proportion of deaths for each entry and adding the a new column in the dataframe with a name of Proportions of death .

clinic\_1 = yearly\_df[yearly\_df["clinic"] == "clinic 1"]

clinic\_2 = yearly\_df[yearly\_df["clinic"] == "clinic 2"]

Seperating the database for clinic1 and clinic 2

fig,ax = plt.subplots(figsize = (10,4))

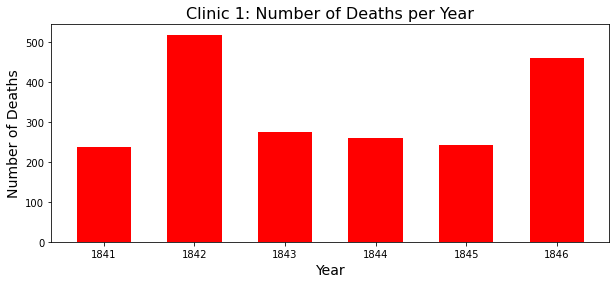
plt.bar(clinic\_1.year, clinic\_1.deaths, width= 0.6, color= "red")

plt.title("Clinic 1: Number of Deaths per Year", fontsize=16)

plt.xlabel("Year", fontsize=14)

plt.ylabel("Number of Deaths", fontsize=14)

Displaying the number of deaths per year for clinic1 using matplotlib



fig,ax = plt.subplots(figsize = (10,4))

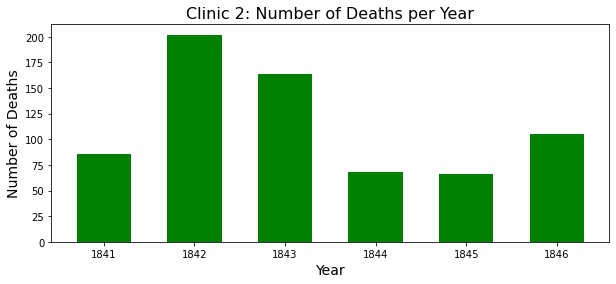
plt.bar(clinic\_2.year, clinic\_2.deaths, width= 0.6, color= "green")

plt.title("Clinic 2: Number of Deaths per Year", fontsize=16)

plt.xlabel("Year", fontsize=14)

plt.ylabel("Number of Deaths", fontsize=14)

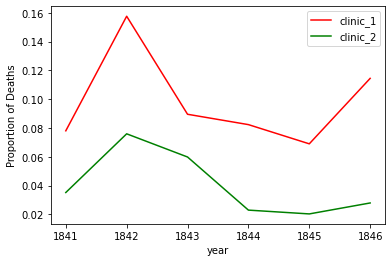
Displaying the number of deaths per year for clinic2 using matplotlib



ax= clinic\_1.plot(x= "year", y= "Proportion of Deaths", label= "clinic\_1", color="red")

clinic\_2.plot(x= "year", y= "Proportion of Deaths", label= "clinic\_2", ax=ax, ylabel= "Proportion of Deaths", color="green")

Creating a line plot using pandas' built-in plotting functionality, with the proportions of deaths plotted over the years for Clinic 1 and Clinic 2.



monthly\_df = pd.read\_csv("monthly\_deaths.csv")

monthly\_df.head(5)

Reading the Monthly Dataset

monthly\_df.info()

Showing the Information of the dataset

monthly\_df["Proportion of Deaths"]= monthly\_df["deaths"] / monthly\_df["births"]

monthly\_df.head(5)

calculating the proportion of deaths for each entry and adding the a new column in the dataframe with a name of Proportions of death .

monthly\_df.dtypes

monthly\_df['date'] =  pd.to\_datetime(monthly\_df['date'])

converts the "date" column to datetime format using the pd.to\_datetime() function. This function converts the "date" column from its current data type

start\_handwashing = pd.to\_datetime('1847-06-01')

before\_washing = monthly\_df[monthly\_df["date"] < start\_handwashing]

after\_washing = monthly\_df[monthly\_df["date"] >= start\_handwashing]

split the DataFrame monthly\_df into two separate DataFrames based on a condition related to the "date" column.

fig,ax = plt.subplots(figsize = (10,4))

x= before\_washing["date"]

y= before\_washing["Proportion of Deaths"]

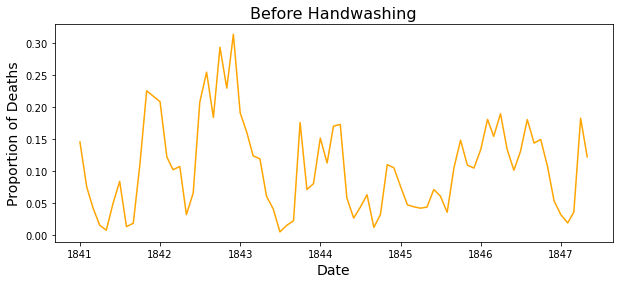
plt.plot(x, y, color= "orange")

plt.title("Before Handwashing", fontsize=16)

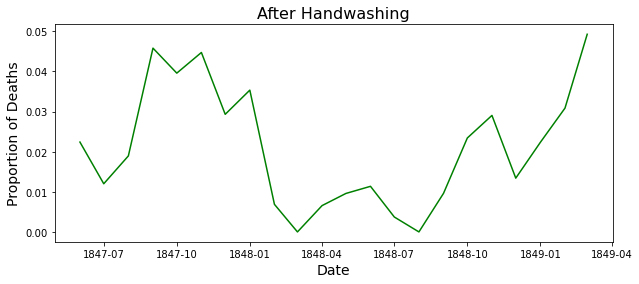
plt.xlabel("Date", fontsize=14)

plt.ylabel("Proportion of Deaths", fontsize=14)

creates a line plot using matplotlib, showing the proportion of deaths over time before the implementation of handwashing.



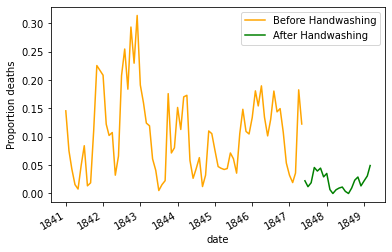
Creating a line plot using matplotlib, showing the proportion of deaths over time after the implementation of handwashing.



ax= before\_washing.plot(x= "date", y= "Proportion of Deaths", label= "Before Handwashing", color="orange")

after\_washing.plot(x= "date", y= "Proportion of Deaths", label= "After Handwashing", ax=ax, ylabel= "Proportion deaths", color="green")

creating a line plot, displaying the proportion of deaths over time before and after the implementation of handwashing on the same graph.



before\_proportion = before\_washing["Proportion of Deaths"]

after\_proportion = after\_washing["Proportion of Deaths"]

before\_proportion.mean()

calculating exactly how much did handwashing decreased the proportion of deaths on average.

after\_proportion.mean()

mean\_diff = after\_proportion.mean() - before\_proportion.mean()

mean\_diff

O/p : -0.0839566075118334

Calculating the difference between both proportions

And here minus sign in the output indicates that there is a disease .